Partnership for Quality Measurement
Powered by Battelle

National Consensus Development and Strategic Planning for Health Care Quality Measurement

Fall 2024 Cycle Endorsement and Maintenance (E&M) Meeting Discussion Guide

ADVANCED ILLNESS AND POST-ACUTE CARE COMMITTEE

February 11, 2025

Prepared by:
Battelle
505 King Avenue, Columbus, Ohio 43201

The analyses upon which this publication is based were performed under Contract Number 75FCMC23C0010, entitled, "National Consensus Development and Strategic Planning for Health Care Quality Measurement," sponsored by the Department of Health and Human Services, Centers for Medicare & Medicaid Services.

Partnership for Quality Measurement
Powered by Battelle

Table of Contents

Overview of Fall 2024 Measures for Review 4
Public Comment 5
Advisory Group Feedback 5
Measures Under Endorsement Review 7
CBE #1623: Bereaved Family Survey - Performance Measure (BFS-PM) Score (%) for all Veteran Affairs Medical Center Inpatient Deaths [Department of Veteran Affairs] 7
Staff Preliminary Assessment Rating 7
Public Comment 9
Advisory Group Feedback 9
Key Themes from Advisory Group Feedback and Staff Assessments 11
CBE #3420: CoreQ: AL Resident Satisfaction Measure [AHCA] 13
Staff Preliminary Assessment Rating 13
Public Comment 15
Advisory Group Feedback 15
Key Themes from Advisory Group Feedback and Staff Assessments 17
CBE #3422: CoreQ: AL Family Satisfaction Measure [AHCA] 19
Staff Preliminary Assessment Rating 19
Public Comment 21
Advisory Group Feedback 21
Key Themes from Advisory Group Feedback and Staff Assessments 21
CBE #4630: Cross-Setting Discharge Function Score for Inpatient Rehabilitation Facilities [RTI International/CMS] 23
Staff Preliminary Assessment Rating 23
Public Comment 24
Advisory Group Feedback 24
Key Themes from Advisory Group Feedback and Staff Assessments 26
CBE #4635: Cross-Setting Discharge Function Score for Long-Term Care Hospitals [RTI International/CMS] 28
Staff Preliminary Assessment Rating 28
Public Comment 29
Advisory Group Feedback 29
Key Themes from Advisory Group Feedback and Staff Assessments 29

CBE #4640: Cross-Setting Discharge Function Score for Skilled Nursing Facilities [RTI International/CMS] 31
Staff Preliminary Assessment Rating 31
Public Comment 32
This measure did not receive any comments during the public comment period. 32
Advisory Group Feedback 32
Key Themes from Advisory Group Feedback and Staff Assessments 33
CBE #4645: Cross-Setting Discharge Function Score – for Home Health Agencies [Abt Global/CMS] 34
Staff Preliminary Assessment Rating 34
Public Comment 35
This measure did not receive any comments during the public comment period. 35
Advisory Group Feedback 35
Key Themes from Advisory Group Feedback and Staff Assessments 35
CBE #3645: Hospice Visits in the Last Days of Life [Abt Global/CMS] 36
Staff Preliminary Assessment Rating 36
Public Comment 37
Advisory Group Feedback 37
Key Themes from Advisory Group Feedback and Staff Assessments 38

List of Tables

Table 1. Overview of Measures Under Endorsement Review 4

List of Figures

Figure 1. Fall 2024 Measures for Committee Review 4
Figure 2. Viewing Public Comments and Developer Responses 5

Overview of Fall 2024 Measures for Review

During this measure review cycle, developers and stewards submitted 8 measures to the Advanced Illness and Post-Acute Care committee for endorsement consideration (Table 1). The measures focused on hospice care, discharge function scores across various care settings, and patient and family satisfaction surveys (Figure 1).

Table 1. Overview of Measures Under Endorsement Review

CBE Number	Measure Title	New/Maintenance	Developer/Steward
1623	Bereaved Family Survey-Performance Measure (BFS-PM) Score (%) for all Veteran Affairs Medical Center Inpatient Deaths	Maintenance	Department of Veteran Affairs
3420	CoreQ: AL Resident Satisfaction Survey	Maintenance	American Health Care Association (AHCA)
3422	CoreQ: AL Family Satisfaction Measure	Maintenance	AHCA
4630	Cross-Setting Discharge Function Score for Inpatient Rehabilitation Facilities	New	RTI International/Centers for Medicare & Medicaid Services (CMS)
4635	Cross-Setting Discharge Function Score for Long-Term Care Hospitals	New	RTI International/CMS
4640	Cross-Setting Discharge Function Score for Skilled Nursing Facilities	New	RTI International/CMS
4645	Cross-Setting Discharge Function Score- for Home Health Agencies	New	Abt Global/CMS
3645	Hospice Visits in the Last Days of Life	Maintenance	Abt Global/CMS

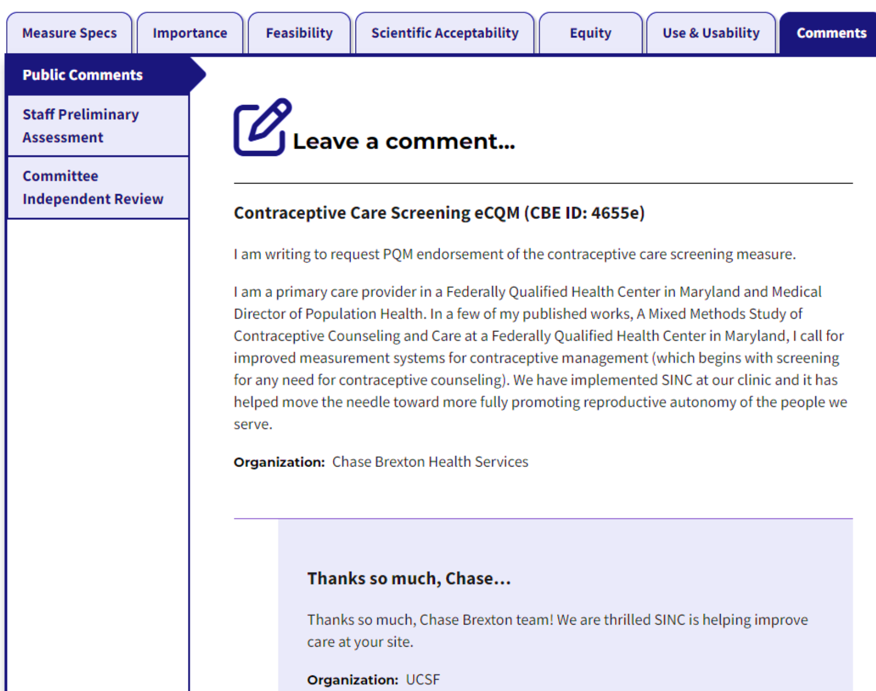
Figure 1. Fall 2024 Measures for Committee Review

Public Comment

Battelle accepts comments on measures under endorsement review through the Partnership for Quality Measurement (PQM) website and Public Comment Listening Sessions. In this evaluation cycle, the public comment period opened on November 15, 2024, and closed on December 16, 2024. Battelle held a Public Comment Listening Session on November 21, 2024.

After the public comment period closed, developers/stewards had the opportunity to respond to public comments on the measure page in the Submission Tool and Repository Measure Database (STAR). To view the public comments and response, go to the "Comments" tab in the left navigation pane (Figure 2). Each comment has a bold heading followed by the body of the comment. Developer responses, if any, appear as a shaded reply beneath the comments. Note that developers are not obligated to respond to public comments. Lastly, the measure evaluation summaries below contain the number of public comments received for each measure.

Figure 2. Viewing Public Comments and Developer Responses



Advisory Group Feedback

The Advisory Group convened on December 4, 2024. In all, 18 of 26 (69%) active Advisory Group members attended to share feedback and ask questions regarding the measures under endorsement review. Developers/stewards of the respective measures also attended and provided responses to the Advisory Group questions. After the meeting, developers/stewards had the opportunity to submit additional written responses to Advisory Group member feedback and questions.

The measure evaluation summaries in this discussion guide contain overviews of the Advisory Group member discussions and developer/steward responses.

To support the review of the public comments and Advisory Group summaries, the number of comments received or number of individuals who shared similar comments, feedback, and/or questions is represented as “a few” (two to three individuals), “several” (four to six individuals), and “many” (more than six individuals). This discussion guide also employs four key categories—Supportive, Dissenting, Mixed, and Probing—to structure and enhance the Recommendation Group discussion.

- **Supportive**: This includes views and comments that express agreement, encouragement, or reinforcement of the measure.
- **Dissenting**: This captures opinions that disagree with or oppose what has been stated about the measure or what has been provided within the measure submission.
- **Mixed**: This category encompasses feedback that contains both supportive and dissenting elements.
- **Probing**: This involves questions or comments that seek to explore, clarify, or delve deeper into aspects of the measure.

Measures Under Endorsement Review

CBE #1623: Bereaved Family Survey - Performance Measure (BFS-PM) Score (%) for all Veteran Affairs Medical Center Inpatient Deaths [Department of Veteran Affairs]

Specifications

Measure Description: *The Bereaved Family Survey-Performance Measure (BFS-PM) is an outcome measure that is used to assess overall quality of care in the last month of life. Currently, the BFS is administered to the next-of-kin of all Veterans who die in a VA inpatient setting (i.e., acute units, intensive care units, inpatient hospice and palliative care units, and VA nursing homes) 4-6 weeks post-death. The BFS-PM is calculated using the global rating item included on the 20-item BFS that has separate versions for male and female Veterans and is available in English and Spanish. The BFS global rating item asks: "Using any number from 0 to 10, where 0 is the worst care possible and 10 is the best care possible, what number would you use to rate the care [he/she] received in the last month of life?" The BFS-PM is calculated as the proportion of family members who provided a "top box" rating of 9 or 10 vs. 0-8 on the global rating item. BFS-PM scores are used for the purposes of monitoring quality of care for Veterans at the end of life nationally, facility benchmarking within the VA healthcare system, and targeting quality improvement efforts.*

Staff Preliminary Assessment Rating[1]

Importance: Met

Rationale: This maintenance measure meets all criteria for "Met" due to its robust evidence base, clear business case, documented performance gap, significant anticipated impact, well-articulated logic model, and its superiority over existing measures, making it essential for addressing quality of care in the last month of life.

Feasibility: Met

Rationale: This established measure meets all criteria for "Met" due to its well-documented feasibility assessment, clear and implementable data collection strategy, and transparent handling of licensing and fees, ensuring practical implementation within the health care system.

Reliability: Met

Rationale: The results demonstrate sufficient reliability at the patient or encounter and accountable-entity levels.

Validity: Not Met, but Addressable

Rationale: The validity testing results support a reasonably strong inference of validity for the measure, confirming that the measure accurately reflects performance on quality or resource use and can distinguish good from poor performance.

---

[1] Located under the "Comments" tab, then "Staff Preliminary Assessment."

The developer conducted statistical risk adjustment, but it is unclear how the developer selected the final risk variables to be included in the model based on the conceptual model and/or overall approach. The developer reported a C-statistic of 0.60, indicating moderate model discrimination.

Equity: Met

Rationale: The measure sufficiently assesses equity in health care delivery and outcomes, providing crucial insights into how different populations are affected by current practices. The methodology and empirical testing are robust, ensuring that the measure can effectively contribute to ongoing efforts to address and reduce health care inequities.

Use & Usability: Met

Rationale: For maintenance, the measure is actively used in at least one accountability application, with a clear feedback approach that allows for continuous updates based on stakeholder feedback. The measure also demonstrates a positive trend in performance results, affirming its ongoing usability. The developer reports no unexpected findings.

Public Comment

Number of Comments Received During the Public Comment Period: 5

Comments and their responses from measure developers can be found on the measure page under the “Comments” tab (*Figure 2*).

Advisory Group Feedback

Feedback/Questions	Summary of Developer Response
Quality Initiative (QI) Improvements: An Advisory Group member asked what has been learned from the measure during its use in QI initiatives.	The BFS is integrated within the Hospice and Palliative Care Program, which is dedicated to implementing quality improvement efforts. They hold phone calls with facilities to make targeted improvements to the quality of end-of-life care based on survey findings.
Variation in Scores: An Advisory Group member requested more information about the variation in scores, as survey scores typically have little variation.	The performance gap table shows a range of 28% to 96%, indicating wide variation in scores nationally. Data can be used to target facilities in need of further support and QI.
Missed Individuals: An Advisory Group member asked if it is possible to miss veterans because of administrative processes or procedures.	They are confident they are not missing veterans as their electronic health record (EHR) allows them to track veterans even if they change units. Through their EHR, they can identify the unit they had care in and then link the quality of care.
Importance: Several Advisory Group members believed this was an important measure. One pointed out that the measure aligns with veterans’ preference for receiving care in the VA. A few expressed that they wished the survey could be translated for use outside of the VA.	They are proud of their survey and viewed the high response rate as a sign of how valuable the survey is to families. In terms of implementing this measure outside of the VA, a handful of health care systems are in various stages of implementation, including Duke Health, Stanford Health Care, UCLA (University of California, Los Angeles) Health, and Kaiser Permanente.
Survey System: An Advisory Group member requested more information about the system used to distribute the survey, given its perceived robustness.	There is a rigorous process to administer the survey. They mail out the survey in 4-6 weeks. If they do not receive a response, they have extensive follow-up, which includes phone calls. Even when they reach the final phone call, veteran families are often interested in speaking.
Response Rate: Several Advisory Group members observed that a 40% response rate for a survey is impressive and indicative of the measure’s significance. One member highlighted that, in comparison, the National Bereavement Survey for hospice patients typically sees a 15-18% response rate. Another member inquired about efforts to further improve the response rate.	They are proud of their response rate, particularly given how infrequently people respond to surveys. They agreed that the response rate reflects the importance and value that veteran families place on the survey. In terms of improving their response rates, they worked with staff members on their results. They highlighted that the staff members also

Feedback/Questions	Summary of Developer Response
	recognized the value of the work and as an opportunity to provide potential resources, such as bereavement contacts, to the families.
Mode of Administration: One Advisory Group member questioned whether adding mail to the existing phone-based survey method impacted the response rate. Another member inquired if the developer had considered using text messages as a strategy to enhance response rates.	When they switched from a phone-only survey to a combination of phone and mail, they noticed a slight decrease in the response rate; previously, when they were only using phone calls, their response rate was around 45%. Although they have not implemented texting yet, they do offer an online QR code. However, only about 5% of their responses are received through this online method, possibly because their target audience primarily consists of individuals over the age of 65.
Value: An Advisory Group member inquired whether any steps had been taken to solicit feedback from family members regarding the value of the survey and its constructs.	In addition to the quantitative items, the survey includes two open-ended questions at the end, designed to allow families to provide narrative feedback. These responses are reviewed for quality improvement and are considered the primary method for engaging with families.
Access to Outpatient Palliative Care Resources: An Advisory Group member asked if a veteran who died with an inpatient status would have access to the outpatient end-of-life caregiving resources that the survey assesses.	The measure looks at the last 30 days of life. They have explored timing of palliative services as a predictor of score and found an association. Therefore, they view receiving those consultations as a process that can improve bereaved scores.
Family Member vs. Person Receiving the Care: An Advisory Group member asked whether there has been any analysis on the potential differences in perceived quality between the family members and the person who received the service, noting that the preferences of the family might differ from those of the patient.	They conducted internal studies between inpatient veteran scores and the results of the survey and found a positive correlation. The literature also states these responses can be largely correlated.
Equity and Race: A few Advisory Group members asked the developer about their equity findings, noting there were disparities at the facility level.	They analyzed measure scores by race and ethnicity and found that Black and African-American families were half as likely to rate their care as excellent compared to their white counterparts. The quantitative items on the survey alone did not explain the differences in care, so they looked closely at the qualitative comments and found that that many of the differences were driven by basic interactions and poor communication. These findings were shared with the Undersecretary of Health. While they do not currently report scores by race or ethnicity at the facility, they are in preliminary discussions to create an annual disparities report to see trends at a national level and identify the facilities in which the largest disparities exist.

Feedback/Questions	Summary of Developer Response
Unexpected Findings: An Advisory Group asked the developer to expand on suicidality among bereaved family members as one of the unexpected findings that have been identified in the process of collecting the BFS.	The developer said that while this situation has only arisen a handful of times when they have been conducting the survey, they said they have developed a system so that if anything alarming is heard or read in survey responses, they ensure the respondent is connected with appropriate mental health resources.
Decision for Top Box Scoring: An Advisory Group member inquired whether any testing was conducted regarding the decision to use a top box score of 9 or 10. They also asked whether the reliability scores or factor loading improved or worsened when a different top box score was used.	The decision to use “top box” ratings of 9 or 10 for the BFS overall rating measure was not based on empirical data. Instead, they dichotomized the scores as 9 or 10 versus 0-8 to align with and facilitate comparisons to quality of end-of-life care ratings in the community, such as those used by the Consumer Assessment of Healthcare Providers and Systems (CAHPS)-Hospice, which also considers 9 or 10 as their “top box” ratings. Consequently, they did not conduct further testing on reliability or factor loadings using different “top box” cutoffs.

Key Themes from Advisory Group Feedback and Staff Assessments

Discussion Category	Key Themes	Source of Comment	Summary of Comments
Supportive	Importance and Value	Advisory Group; Public Comment	The Advisory Group discussed how this is an important survey that seems to be valuable to veterans and their families, which can be seen through high response rates, and expressed interest in seeing the measure transition to outside of the VA. Public comments highlighted the measure’s significance and expressed interest in extending its application to non-VA settings to encompass a wider range of end-of-life care experiences.
	High Response Rates	Advisory Group	The Advisory Group praised the high response rates and the rigorous system the developer has put into place to ensure those response rates. They discussed that the response rates could also represent the importance and value of the measure. They also discussed potential ways the response rates could improve even further.
	QI Processes	Advisory Group	The Advisory Group discussed the way this measure drives QI processes, including the ways in which it could drive potential equity considerations, particularly surrounding race and ethnicity.
Dissenting	Risk Adjustment	Staff Assessment	The developer conducted statistical risk adjustment, but it is unclear how the developer selected the final risk variables to be included in the model based on the conceptual model and/or overall approach. The developer reported a C-statistic of 0.60, indicating moderate model discrimination.

Discussion Category	Key Themes	Source of Comment	Summary of Comments
Probing	Race and Equity	Advisory Group	The Advisory Group was interested in hearing more about how this measure could drive equity issues, particularly related to race.
	Unexpected Findings	Advisory Group	The Advisory Group asked for more information on how the developer contends with bereaved family members such as those who express suicidality and require additional mental health resources.

CBE #3420: CoreQ: AL Resident Satisfaction Measure [AHCA]

Specifications

Measure Description: *The measure calculates the percentage of Assisted living (AL) residents, those living in the facility for two weeks or more, who are satisfied. This patient reported outcome measure is based on the CoreQ: AL Resident Satisfaction questionnaire that is a four-item questionnaire.*

Staff Preliminary Assessment Rating

Importance: Not Met but Addressable

Rationale: This previously endorsed measure meets many criteria for "Met" due to its clear business case, documented performance gap, and its well-articulated logic model. Measuring resident satisfaction is essential to helping AL facilities understand patient preferences, supporting patients and their families in choosing facilities, and allowing AL facilities to monitor and improve the quality of care they provide. However, the Recommendation Group should consider if more recent evidence should be incorporated into the literature review.

Feasibility: Not Met but Addressable

Rationale: This previously endorsed measure meets most criteria for "Met" due to its established data collection strategy and lack of licensing and fees, ensuring practical implementation within the health care system. However, it is unclear if all data used to calculate the measure is in an electronic format, and if not, if there is a near-term plan to support routine and electronic data capture.

Reliability: Not Met but Addressable

Rationale: The current accountable entity-level reliability metrics do not meet the established thresholds, indicating potential issues with the consistency and accuracy of the results across different settings and populations. However, the identified limitations are deemed addressable, as the developer may consider increasing the sample size to meet the requirements of the selected statistical methods. By addressing this/these issue/s, there is potential to enhance the reliability.

Validity: Not Met but Addressable

Rationale: The data should be updated to reflect more recent performance. Going forward, additional studies that either rule out potential confounding or describe features of potential mechanisms will strengthen causal claims.

Measure is not risk adjusted. Rationale for not performing risk adjustment is based on dated literature (2003-2014) and should be reassessed.

Equity: Not Met but Addressable

Rationale: While the measure attempts to assess equity in health care delivery and outcomes, additional work is needed to ensure the measure provides valid comparisons between Black and white residents who responded to the CoreQ: AL Resident Satisfaction Questionnaire. This

limits the ability to provide a comprehensive understanding of the differences in performance across different populations.

Use & Usability: Met

Rationale: For maintenance, the measure is actively used in at least one accountability application, with a clear feedback approach that allows for continuous updates based on stakeholder feedback. The developer reports no unexpected findings. Usability data demonstrate the measure is sensitive to changes in the care environment.

Public Comment

Number of Comments Received During the Public Comment Period: 1

Comments and their responses from measure developers can be found on the measure page under the “Comments” tab (*Figure 2*).

Advisory Group Feedback

Feedback/Questions	Summary of Developer Response
Equity: An Advisory Group member asked if, outside of race, certain populations may be over- or underrepresented, especially related to income.	Over the past 2 years, they have continually looked at new variables to include in the survey, including a socioeconomic status variable, insurance, and age. They also have been looking at insurance data as a proxy for income.
Target Population: An Advisory Group member asked about adults who may reside in assisted living but fall outside of the age range, such as those who have had an injury or stroke.	The minimum representative age for the assisted living population is predominately 65. The number of individuals who would be in assisted living under the age of 65 would be small, regardless of injury or condition. Efforts are currently underway to improve the capture of data concerning individuals who are receiving memory care as part of their assisted living services.
Decision for Top Box Scoring: An Advisory Group member asked if any testing was conducted related to the decision to top box at 9 or 10. They also asked if reliability scores or factor loading improved or worsened with a different top box.	They tested multiple scales and found that the industry prefers the current scale. The intent of the CoreQ is to enable facilities to benchmark themselves against others. Even if the scoring system is changed and the distribution of scores is flattened, the relative standing of the facilities tends not to change significantly.
Race: An Advisory Group member noted the submission shows that Black residents tend to be in lower-performing facilities. They questioned whether the developer had initiated contact with these facilities to discuss potential improvements. Another Advisory Group member emphasized the importance of this measure, noting that it could significantly aid facilities, particularly those serving minority patients, by incorporating direct feedback from families and residents to drive improvements	For the last 2 years, they have focused on reaching out to facilities with higher minority populations, particularly Black residents. They acknowledged the low number of minority respondents in their surveys, attributing this partly to the demographic composition of assisted living facilities rather than to response bias. To facilitate this outreach and improve participation, surveys are conducted at no cost in these selected facilities.
Meaningfulness to Residents: An Advisory Group member commented that the survey items were selected based on a focus group of 40 residents some years ago. They asked if meaningfulness has been examined since then.	Although no focus groups have been conducted with the facilities themselves since the development of the CoreQ, they have engaged in focus groups with caregivers, certified nurse aides, administrators, family members, residents, and other survey vendors.
Exclusion Criteria: An Advisory Group member observed that residents receiving hospice services are excluded from the measure and questioned whether the developer had considered also excluding	They had not previously been asked about excluding home health services from external organizations and would need to discuss with their team to consider and collect data.

Feedback/Questions	Summary of Developer Response
home health services provided by external organizations. They highlighted the difficulty in differentiating between various services when they are delivered by external providers.	
Response Rate: An Advisory Group member asked about the survey response rate and how many facilities participate.	The average CoreQ response rate is 64%, which is reflective of the survey's brevity and ease of completion. The response rate is even higher for residents (CBE #3422) at approximately 90%. About 400 facilities complete the survey.
Distribution: An Advisory Group member asked how the CoreQ is distributed to assisted living residents.	Packets are sent to the facilities to distribute, with one survey per resident. Residents must complete the CoreQ on their own; If the respondent answers otherwise, their survey is excluded. Residents of assisted living tend to be healthier than the standard population in a nursing home. As most facilities are requesting the surveys, they are unlikely to interfere with the data.
Age of Evidence: An Advisory Group member observed that the literature review seems outdated and inquired about the frequency of publications for assisted living settings compared to nursing homes and other settings. They asked the Recommendation Group to consider the age of the literature review, and a potential update, during their discussion.	While much of their literature is dated, the content has not changed significantly. Assisted living data is not as standardized as nursing home data and results in fewer available repositories of secondary data, necessitating reliance on primary data collection for research and analysis in assisted living contexts. This situation contributes to the difficulties in updating and expanding the literature base for assisted living.
Patient-Centered Care: An Advisory Group member noted that satisfaction appears to be related to patient-centered care, although an individual's satisfaction might be influenced by a lack of comparison if they have not experienced alternative care options.	Assisted living setting have fewer standardized surveys compared to other settings. The CoreQ is short and allows the facilities to add other items to the bottom of the survey to gather additional information.
Timeframe: An Advisory Group member asked when the surveys are administered.	They provide guidance on survey timing to the facilities, emphasizing the importance of consistency. While some developers distribute the survey biannually, most opt for an annual schedule. They recommend avoiding survey distribution during holidays or staff vacations to ensure better response rates. Respondents are given 2 months to complete the survey, and residents who have been at the facility for less than 2 weeks are excluded from participation. They monitor for fluctuations in score and, for the most part, the scores are stable.
Measure Design: An Advisory Group member asked how the residents and families of residents are included in the design of the measure and analysis of the results.	Families and residents were involved in selecting the items they believed were most important and in the wording of the items themselves. The developer performed cognitive testing to ensure that respondents understood what was being asked. They also involved families, caregivers, and residents in the selection of the response scale, formatting of the questionnaire, and wording of the introductory

Feedback/Questions	Summary of Developer Response
	letter. However, these groups are minimally involved in the analysis process.
Truly Reflective of "Satisfaction": An Advisory Group member stated that ratings on specific elements do not necessarily equate to overall satisfaction. Another Advisory Group member questioned why the developer didn't simplify to a single question: "Are you satisfied with your care?"	They clarified that while they call the CoreQ a "satisfaction survey," it is more accurate to call it a "rating survey." They initially tested a longer, comprehensive survey with multiple items to assess its correlation with a shorter version. With CoreQ, they aimed to include distinct, non-overlapping items. They tested direct satisfaction questions but found that the item pertaining to recommending the facility to family and friends had better results.

Key Themes from Advisory Group Feedback and Staff Assessments

Discussion Category	Key Themes	Source of Comment	Summary of Comments
Dissenting	Evidence	Advisory Group; Staff Assessment	An Advisory Group requested the Recommendation Group keep the age of the evidence in mind and consider whether an update would be appropriate. The staff assessment also identified that that nearly all literature cited is from before 2018 and may not reflect recent advances in this area.
	Feasibility	Staff Assessment	Facilities had all information needed available except for cognitive status, which is necessary to determine residents whose surveys should be excluded from the measure. It is unclear if all data used to calculate the measure is in an electronic format, and if not, if there is a near-term plan to support routine and electronic data capture.
	Reliability	Staff Assessment	The developer appears to have conducted a bootstrap version of reliability at the accountable entity-level. Only the mean signal-to-noise reliability (0.84) is given so there is insufficient evidence to know whether >70% of entities have reliability >0.60.
	Validity	Staff Assessment	The developer conducted validity testing using data from 2018. Additional studies that either rule-out potential confounding or describe features of potential mechanisms will strengthen causal claims. In addition, the rationale for not performing risk adjustment is based on dated literature (2003-2014) and should be reassessed.
Probing	Satisfaction	Advisory Group	The Advisory Group discussed whether the measure truly captures the concept of "satisfaction" and how the measure could still potentially play a role in patient-centered care.

Discussion Category	Key Themes	Source of Comment	Summary of Comments
	Equity	Advisory Group; Staff Assessment	The Advisory Group discussed how the measure deals with issues of equity, particularly race, as Black and African-American residents tend to live in lower-performing assisted living facilities. A few members believed this measure could help drive better performance in those facilities. They also touched upon other variables, such as age and income status. The staff assessment also identified that additional work is needed to ensure the measure provides valid comparisons between Black and white residents who responded to the CoreQ: AL Resident Satisfaction Questionnaire.

CBE #3422: CoreQ: AL Family Satisfaction Measure [AHCA]

Specifications

Measure Description: *The measure calculates the percentage of family or designated responsible party for assisted living (AL) residents who are satisfied. This consumer reported outcome measure is based on the CoreQ: AL Family Satisfaction questionnaire that has three items.*

Staff Preliminary Assessment Rating

Importance: Not Met but Addressable

Rationale: This previously endorsed measure meets many criteria for “Met” due to its clear business case, documented performance gap, and its well-articulated logic model. Measuring family satisfaction is essential to helping AL facilities understand patient preferences, supporting patients and their families in choosing facilities, and allowing AL facilities to monitor and improve the quality of care they provide. However, the committee should consider if more recent evidence should be incorporated into the literature review. Further, the performance gap decile data submitted for this measure are identical to those submitted for measure #3420. The committee may wish to seek clarification on this overlap.

Feasibility: Not Met but Addressable

Rationale: This previously endorsed measure meets most criteria for “Met” due to its established data collection strategy and lack of licensing and fees, ensuring practical implementation within the health care system. However, it is unclear if all data used to calculate the measure is in an electronic format, and if not, if there is a near-term plan to support routine and electronic data capture.

Reliability: Not Met but Addressable

Rationale: The current accountable entity-level reliability metrics do not meet the established thresholds, indicating potential issues with the consistency and accuracy of the results across different settings and populations. However, the identified limitations are deemed addressable, as the developer may consider increasing the sample size to meet the requirements of the selected statistical methods. By addressing this/these issue/s, there is potential to enhance the reliability.

Validity: Not Met but Addressable

Rationale: The data should be updated to reflect more recent performance. Going forward, additional studies that either rule out potential confounding or describe features of potential mechanisms will strengthen causal claims.

Measure is not risk adjusted. Rationale for not performing risk adjustment is based on dated literature (2003-2014) and should be reassessed.

Equity: Not Met but Addressable

Rationale: While the measure attempts to assess equity in health care delivery and outcomes, additional work is needed to ensure the measure provides valid comparisons between the family

members of Black and white residents who responded to the CoreQ: AL Resident Satisfaction Questionnaire. This limits the ability to provide a comprehensive understanding of the differences in performance across different populations.

Use & Usability: Met

Rationale: For maintenance, the measure is actively used in at least one accountability application, with a clear feedback approach that allows for continuous updates based on stakeholder feedback. The developer reports no unexpected findings. Usability data demonstrate the measure is sensitive to changes in the care environment.

Public Comment

Number of Comments Received During the Public Comment Period: 1

Comments and their responses from measure developers can be found on the measure page under the "Comments" tab (*Figure 2*).

Advisory Group Feedback

Feedback/Questions	Summary of Developer Response
Similar Feedback to CBE #3420: Advisory Group members had similar questions and feedback surrounding Decision for Top Box Scoring, Exclusion Criteria, Response Rate, Patient-Centered Care, Timeframe, Meaningfulness to Residents, Measure Design, Single Question, Age of Evidence, and Truly Reflective of "Satisfaction" as noted within CBE #3420.	Please refer to the developer's responses for CBE #3420.
Validity: An Advisory Group member asked the developer to explain the low correlation between the family's CoreQ score and the measures used to establish construct validity.	While low correlation could be considered a failing, it could also mean the measure is picking up on a different aspect of quality.

Key Themes from Advisory Group Feedback and Staff Assessments

Discussion Category	Key Themes	Source of Comment	Summary of Comments
Dissenting	Evidence	Advisory Group; Staff Assessment	An Advisory Group requested the Recommendation Group keep the age of the evidence in mind and whether an update would be appropriate. The staff assessment also identified that the primary rationale for the impact of the measure is a 2014 systematic review showing a positive effect of patient-provider relationships on health care outcomes and a 2013 systematic review that found person-centered care was associated with psychosocial benefits. However, the Recommendation Group should consider if more recent evidence should be incorporated into the literature review.
	Feasibility	Staff Assessment	Facilities had all information needed available except for cognitive status, which is necessary to determine residents whose surveys should be excluded from the measure. It is unclear if all data used to calculate the measure is in an electronic format, and if not, if there is a near-term plan to support routine and electronic data capture.

Discussion Category	Key Themes	Source of Comment	Summary of Comments
	Reliability	Staff Assessment	The developer appears to have conducted a bootstrap version of reliability at the accountable entity-level. Only the mean signal-to-noise reliability (0.82) is given so there is insufficient evidence to know whether >70% of entities have reliability >0.60.
	Validity	Staff Assessment	The developer conducted validity testing using data from 2018. Additional studies that either rule-out potential confounding or describe features of potential mechanisms will strengthen causal claims. In addition, the rationale for not performing risk adjustment is based on dated literature (2003-2014) and should be reassessed.
Probing	Satisfaction	Advisory Group	The Advisory Group discussed whether the measure truly captures the concept of “satisfaction” and how the measure could still potentially play a role in patient-centered care.
	Equity	Staff Assessment	For this optional domain, the staff assessment identified that additional work is needed to ensure the measure provides valid comparisons between the family members of Black and white residents who responded to the CoreQ: AL Family Satisfaction Questionnaire.

CBE #4630: Cross-Setting Discharge Function Score for Inpatient Rehabilitation Facilities [RTI International/CMS]

Specifications

Measure Description: *This outcome measure estimates the percentage of Inpatient Rehabilitation Facility (IRF) Medicare patient stays that meet or exceed an expected discharge function score. The expected discharge function score is a risk-adjusted estimate that accounts for patient characteristics. The measure includes patients who are 18 years of age or older and the timeframe for the measure is 12 months.*

Staff Preliminary Assessment Rating

Importance: Met

Rationale: This new measure meets all criteria for “Met” due to its evidence base, clear business case, documented performance gap, and significant anticipated impact.

Feasibility: Met

Rationale: This new measure meets all criteria for “Met” due to its feasibility assessment, clear and implementable data collection strategy, and transparent handling of licensing and fees, ensuring practical implementation within the health care system.

Reliability: Met

Rationale: The results demonstrate sufficient reliability at the patient or encounter and accountable-entity levels.

Validity: Met

Rationale: As a new measure, person- or episode-level validity assessment is sufficient. Going forward, additional studies that either rule out potential confounding (which includes the risk-adjustment model) or describe features of potential mechanisms will strengthen causal claims.

The risk adjustment methods used are appropriate and demonstrate variation in the prevalence of risk factors across measured entities, contribute to unique variation in the outcome, and show the impact of risk adjustment for providers at high or low extremes of risk. The model performance is acceptable.

Equity: Met

Rationale: The measure sufficiently assesses equity in health care delivery and outcomes, providing crucial insights into how different populations are affected by current practices. The methodology and empirical testing are robust, ensuring that the measure can effectively contribute to ongoing efforts to address and reduce health care inequities.

Use & Usability: Met

Rationale: The measure is actively used in at least one accountability application, with a clear feedback approach that allows for continuous updates based on stakeholder feedback. The specific actions accountable entities can take to improve measure results are described within the logic model and evidence of measure importance.

Public Comment

Number of Comments Received During the Public Comment Period: 12

Comments and their responses from measure developers can be found on the measure page under the “Comments” tab (*Figure 2*).

Advisory Group Feedback

Feedback/Questions	Summary of Developer Response
Gaming: A few Advisory Group members asked for more information about how the expected discharge function score is calculated or set, expressing concern that it could be manipulated so entities could appear to be doing better on the measure. An Advisory Group member asked for clarification on what data are being gathered by clinicians.	The expected discharge function scores are not set by the clinicians; rather, they are calculated by CMS and CMS’s contractors using national Inpatient Rehabilitation Facility Patient Assessment Instrument (IRF-PAI) data. The other Discharge Function Score measures are calculated similarly per their respective setting. The calculated score considers the patient’s prior function, comorbidities, and other patient-level factor and characteristics. CMS added that it would be difficult to game a complex mixture of patient characteristics and comorbidities. CMS noted that they have tried to address gaming by using an imputation model to estimate scores for data that are unavailable and by using all the patient characteristics that are available in the assessments. Clinicians are collecting patient characteristics, such as the ability to walk or use a wheelchair, ability to eat, and manage oral hygiene. The observed discharge function score is the sum of the individual function activities at discharge.
Overlapping Measures: A few Advisory Group members asked how this measure is different from measures already applied in the post-acute care setting, particularly those that pertain to self-care and mobility. They asked if the intent of the cross-setting measure is to replace currently in-place measures; if not, they asked what the feasibility of the cross-setting measures is.	The Improving Medicare Post-Acute Care Transformation (IMPACT) Act, passed in 2014, requires CMS to implement a measure across all settings. CMS first implemented a process measure, which is now topped out. CMS is replacing that with these four aligned cross-setting measures. The approach, numerator, and denominator are calculated the same for each measure, but the coefficients and risk adjustments vary by setting and use the data specific to the setting (e.g., the IRF measure is calculated using IRF data). IRF data was not tested against Skilled Nursing Facility (SNF) data for the convergent validity testing. This is a motor function measure that incorporates some self-care and mobility, while other self-care and mobility measures already in place in the IRF and SNF settings have a broader range. For example, the self-care measure in the IRF setting also includes activities related to dressing the upper body, dressing the

Feedback/Questions	Summary of Developer Response
	lower body, and footwear. These three dressing items are not collected in all settings and therefore are not included in the cross-setting measure. CMS also added that they are constantly monitoring measures to observe when they may be topped out or overlap with one another. They highlighted that the Measure Set Review (MSR) process is used to garner recommendations on measures to remove.
Importance: A few Advisory Group members were pleased to see the measures coming forward, and that cross-setting measures are important, particularly because of how difficult it is to integrate and unify models in a meaningful way across settings. Another complimented the information pertaining to risk adjustment and the involvement of patients and technical expert panels (TEPs).	Not applicable.
Equity: An Advisory Group member asked how the developer will continue to evaluate which social risk factors will need to be adjusted for in the measures.	They did a lot of testing for social risk factors. For this measure, social risk factors have a limited impact because patients are staying at a facility, so everything is within control of the facility. Therefore, they did not want to risk adjust many social risk factors and mask actual disparities in care. CMS is interested in making sure that providers are aware of the disparities results.
Use and Usability: An Advisory Group member asked several questions regarding use and usability. A few members of the group echoed that these concepts were important. • How would an IRF use this measure? Should they be using it to compare themselves to other IRFs? • Is the intent for consumers to be able to see the measure scores and compare across the different settings? • Has there been any feedback from providers (even anecdotal) about how they are using this measure for internal quality improvement?	The measures are all calculated based on data for their respective setting. The data are publicly reported on Care Compare, so that people can compare IRFs to IRFs; the intent is to compare setting to setting but not across settings (e.g., not IRFs to SNFs). It is important to keep the settings and populations distinct, as conditions and phase of recovery vary across the settings. They believed the measure information, as presented, was comprehensible to an average person. In a previous analysis, they found that consumers better understood data measuring the percentage of patients who met or exceeded an expected score compared to a measure that looked at a change in function score.
Title: A few Advisory Group members commented that the word “cross-setting” was misleading in the title and could lead to confusion that the measure data were meant to be compared with one another rather than kept separately.	The word “cross-setting” is meant to align the interpretation and methodology across the measures. They reiterated that the settings are meant to remain distinct.
Reports: An Advisory Group member asked when the IRFs began receiving their confidential feedback reports.	Data were reported to providers and public reporting began recently (approximately September 2024 for IRF data).

Feedback/Questions	Summary of Developer Response
Validity: An Advisory Group member pointed out that the risk adjustment relies on comorbidities presented at the time of admission. They asked what the validity of this approach is as well as the validity of not factoring in patients who deteriorate.	IRFs have strict requirements about who they admit, including that the patients need to be able to tolerate 3 hours of therapy per day or 15 hours per week. Therefore, the patients are expected to have meaningful improvement. Patients who are discharged to hospice, individuals who have a medical emergency, and individuals who have an incomplete stay are excluded. Many of the admission factors considered are functions at admission, and those are determined using codes from IRF-PAI data.
Target Population: An Advisory Group member pointed out that 83% of the measure's cohort are traditional Medicare patients. They asked if Medicare Advantage (MA) patients are also included.	The IRF measure includes both MA and Medicare fee-for-service (FFS) patients. They did not test or risk adjust for care source, as they expected the outcomes would not be different. However, they could look into whether there are any disparities.
Unintended Consequences: An Advisory Group member asked how unintended consequences are being monitored for.	One of their primary focuses was not to cause restricted access, which is why the measure has many exclusions.
Quality Initiatives: An Advisory Group member asked if clinicians have provided feedback on if and how QI activities are using this data.	The developer did not specifically address this comment.
Public Comments: An Advisory Group member asked the developer if they could speak to Troy Hillman's public comment made during the November 21 listening session (which mentions topics such as competing measures, concerns about feasibility and the imputation methodology, and unintended consequences).	The developer did not specifically address this comment. Staff note: A developer reply to this public comment can be found under the "comments" tab of the measure page.

Key Themes from Advisory Group Feedback and Staff Assessments

Discussion Category	Key Themes	Source of Comment	Summary of Comments
Supportive	Importance	Advisory Group	A few Advisory Group members expressed that they were pleased to see these measures brought forward and that their cross-setting nature would be beneficial.
Dissenting	Exclusion Criteria	Advisory Group	The Advisory Group discussed whether it is appropriate for the measure to exclude patients who have less certain trajectories and less likelihood to significantly improve their measure scores.
	Overlapping Measures	Advisory Group; Public Comment; Staff Assessment	The Advisory Group discussed whether the new cross-setting measures would overlap with already existing measures.

Discussion Category	Key Themes	Source of Comment	Summary of Comments
			Public commenters highlight inconsistencies with existing endorsed measures, which utilize a broader range of functional items and provide a more accurate representation of patient outcomes. The staff assessment noted it is unclear if the proposed measure overlaps with two existing measures (CBE #2635: Inpatient Rehabilitation Facility [IRF] Functional Outcome Measure: Discharge Mobility Score for Medical Rehabilitation Patients and CBE #2636: IRF Functional Outcome Measure: Discharge Self-Care Score for Medical Rehabilitation Patients).
	Feasibility	Staff Assessment, Public Comment	Public commenters argue that the measure is administratively challenging due to its complex imputation methodology, which requires substantial investment in technology and training. The staff assessment noted, in considering the people necessary to implement the measure, the Recommendation Group should assess whether the imputation methodology used in the measure calculation places an undue burden on providers.
	Unintended Consequence of Patient Access	Public Comment	Public commenters criticized the measure's designation as "cross-setting" due to differing expectations across post-acute care settings, potentially leading to inequitable patient access and misdirected referrals.
Probing	Gaming	Advisory Group	A few Advisory Group members discussed whether the function score could be easily manipulated so that entities could appear to perform better on the measure.

CBE #4635: Cross-Setting Discharge Function Score for Long-Term Care Hospitals [RTI International/CMS]

Specifications

Measure Description: *This outcome measure estimates the percentage of Long-Term Care Hospital (LTCH) patient stays that meet or exceed an expected discharge function score. The expected discharge function score is a risk-adjusted estimate that accounts for resident characteristics. The measure includes patients 18 years of age or older and the measure timeframe is 12 months.*

Staff Preliminary Assessment Rating

Importance: Met

Rationale: This new measure meets all criteria for “Met” due to its evidence base, clear business case, documented performance gap, and significant anticipated impact.

Feasibility: Met

Rationale: This new measure meets all criteria for “Met” due to its feasibility assessment, clear and implementable data collection strategy, and transparent handling of licensing and fees, ensuring practical implementation within the health care system.

Reliability: Met

Rationale: The results demonstrate sufficient reliability at the patient or encounter and accountable-entity levels.

Validity: Met

Rationale: As a new measure, person- or episode-level validity assessment is sufficient. Going forward, additional studies that either rule out potential confounding (which includes the risk-adjustment model) or describe features of potential mechanisms will strengthen causal claims.

The risk adjustment methods used are appropriate and demonstrate variation in the prevalence of risk factors across measured entities, contribute to unique variation in the outcome, and show the impact of risk adjustment for providers at high or low extremes of risk. The model performance is acceptable.

Equity: Met

Rationale: The measure sufficiently assesses equity in health care delivery and outcomes, providing crucial insights into how different populations are affected by current practices. The methodology and empirical testing are robust, ensuring that the measure can effectively contribute to ongoing efforts to address and reduce health care inequities.

Use & Usability: Met

Rationale: The measure is actively used in at least one accountability application, with a clear feedback approach that allows for continuous updates based on stakeholder feedback. The specific actions accountable entities can take to improve measure results are described within the logic model and evidence of measure importance.

Public Comment

This measure did not receive any comments during the public comment period.

Advisory Group Feedback

Feedback/Questions	Summary of Developer Response
Similar Feedback to CBE #4630: Advisory Group members had similar questions and feedback surrounding gaming, overlapping measures, equity, importance, use and usability, and title as noted within CBE #4630.	Please refer to the developer's responses for CBE #4630.
Entity Accountability: A few Advisory Group members pointed out that LTCHs will sometime discharge a patient, particularly underrepresented minorities, too soon and without making sure they are adequately prepared if insurance coverage becomes an issue. The Advisory Group members asked how the measure takes these disparities into consideration and how they might hold an entity more accountable to prevent these premature discharges.	The measure only reports on patients who had a complete stay, regardless of where they were discharged to. They noted that LTCH patients tend to be medically complex, so the discharge-to-community is low. In their analyses, they saw disparities related to race and ethnicity; they noted that they are not adjusting for those so that the disparities will not be masked. CMS also added that this is an outcome-based measure. While it does not look at timeframes, they noted that an entity would be at a disadvantage if they discharged someone prematurely because they would risk receiving a lower score on the measure.

Key Themes from Advisory Group Feedback and Staff Assessments

Discussion Category	Key Themes	Source of Comment	Summary of Comments
Supportive	Importance	Advisory Group	A few Advisory Group members expressed that they were pleased to see these measures brought forward and that their cross-setting nature would be beneficial.
Dissenting	Feasibility	Staff Assessment	In considering the people necessary to implement the measure, the Recommendation Group should assess whether the imputation methodology used in the measure calculation places an undue burden on providers.
	Overlapping Measures	Advisory Group; Staff Assessment	The Advisory Group discussed whether the new cross-setting measures would overlap with already existing measures.

Discussion Category	Key Themes	Source of Comment	Summary of Comments
			The staff assessment noted it is unclear if the proposed measure overlaps with the existing measure (CBE #2632: Long-Term Care Hospital Functional Outcome Measure: Change in Mobility Among Patients Requiring Ventilator Support).
Probing	Gaming	Advisory Group	A few Advisory Group members discussed whether the function score could be easily manipulated so that entities could appear to perform better on the measure.
	Accountability	Advisory Group	The Advisory Group discussed how the measure could play a role in helping prevent premature discharges from LTCHs, particularly as they relate to underrepresented minorities.

CBE #4640: Cross-Setting Discharge Function Score for Skilled Nursing Facilities [RTI International/CMS]

Specifications

Measure Description: *This outcome measure estimates the percentage of Medicare Part A skilled nursing facility stays that meet or exceed an expected discharge function score. The expected discharge function score is a risk-adjusted estimate that accounts for resident characteristics. The measure includes patients who are 18 years of age or older and the measure timeframe is 12 months.*

Staff Preliminary Assessment Rating

Importance: Met

Rationale: This new measure meets all criteria for "Met" due to its evidence base, clear business case, documented performance gap, and significant anticipated impact.

Feasibility: Met

Rationale: This new measure meets all criteria for "Met" due to its feasibility assessment, clear and implementable data collection strategy, and transparent handling of licensing and fees, ensuring practical implementation within the health care system.

Reliability: Met

Rationale: The results demonstrate sufficient reliability at the patient or encounter and accountable-entity levels.

Validity: Met

Rationale: As a new measure, person- or episode-level validity assessment is sufficient. Going forward, additional studies that either rule out potential confounding (which includes the risk-adjustment model) or describe features of potential mechanisms will strengthen causal claims.

The risk adjustment methods used are appropriate and demonstrate variation in the prevalence of risk factors across measured entities, contribute to unique variation in the outcome, and show the impact of risk adjustment for providers at high or low extremes of risk. The model performance is acceptable.

Equity: Met

Rationale: The measure sufficiently assesses equity in health care delivery and outcomes, providing crucial insights into how different populations are affected by current practices. The methodology and empirical testing are robust, ensuring that the measure can effectively contribute to ongoing efforts to address and reduce health care inequities.

Use & Usability: Met

Rationale: The measure is actively used in at least one accountability application, with a clear feedback approach that allows for continuous updates based on stakeholder feedback. The specific actions accountable entities can take to improve measure results are described within the logic model and evidence of measure importance.

Public Comment

This measure did not receive any comments during the public comment period.

Advisory Group Feedback

Feedback/Questions	Summary of Developer Response
Similar Feedback to CBE #4630: Advisory Group members had similar questions and feedback surrounding gaming, importance, equity, use and usability, and title as noted within CBE #4630.	Please refer to the developer's responses for CBE #4630.
Validity: A few Advisory Group members commented on how the convergent testing had the most variability out of the four cross-setting measures.	This measure showed the highest correlation with discharge self-care and mobility. They analyzed individual item data and found that higher scores on each item also correlated with higher discharge to community scores, which supports the validity of the items at discharge.
Target Population: An Advisory Group member asked for clarification on what Medicare populations the measure covers. Upon hearing that the measure only covers Medicare fee-for-service (FFS) patients, several Advisory Group members expressed concerns about the small numbers of patients this measure would then include, highlighting that many SNF patients have Medicare managed care. One said that this could negatively impact the facility's score. Another Advisory Group member asked what the endorsement process would look like if the measure were to eventually expand to Medicare Advantage (MA) patients.	The SNF measure is restricted to FFS because that is the population currently required to submit data. CMS recently held listening sessions regarding this topic. CMS added that they are currently limited by the IMPACT Act but are in the process of working to include other payers in this measure and other settings. Battelle clarified that if the measure population were to expand, the developer would be asked to submit additional information and return for endorsement review.
Dual Eligibility: An Advisory Group member asked for clarification on how this measure affects the dual-eligible population.	Individuals who are dually eligible are included in this measure and all the other settings. They performed a disparity analysis looking at dual-eligible individuals versus Medicare-only individuals. For SNF, they found that facilities with lower percentages of dual-eligible patients tended to do better. They did not adjust for this, as they did not want to mask disparities.

Key Themes from Advisory Group Feedback and Staff Assessments

Discussion Category	Key Themes	Source of Comment	Summary of Comments
Supportive	Importance	Advisory Group	A few Advisory Group members expressed that they were pleased to see these measures brought forward and that their cross-setting nature would be beneficial.
Dissenting	Target Population	Advisory Group	Several Advisory Group members expressed concern over this measure only covering FFS, pointing out that those individuals tend to be a small percentage of patients in SNFs, meaning the measure would have limited applicability and usefulness.
	Feasibility	Staff Assessment	In considering the people necessary to implement the measure, the Recommendation Group should assess whether the imputation methodology used in the measure calculation places an undue burden on providers.
Probing	Gaming	Advisory Group	A few Advisory Group members discussed whether the function score could be easily manipulated so that entities could appear to perform better on the measure.

CBE #4645: Cross-Setting Discharge Function Score – for Home Health Agencies [Abt Global/CMS]

Specifications

Measure Description: *This outcome measure estimates the percentage of Home Health (HH) Medicare patients (18+) who meet or exceed an expected discharge function score over a 12-month period.*

The expected discharge function score is a risk-adjusted estimate that accounts for patient characteristics.

Staff Preliminary Assessment Rating

Importance: Met

Rationale: This new measure meets all criteria for “Met” due to its evidence base, clear business case, documented performance gap, and significant anticipated impact.

Feasibility: Met

Rationale: This new measure meets all criteria for “Met” due to its feasibility assessment, clear and implementable data collection strategy, and transparent handling of licensing and fees, ensuring practical implementation within the health care system.

Reliability: Met

Rationale: The results demonstrate sufficient reliability at the patient or encounter and accountable-entity levels.

Validity: Met

Rationale: As a new measure, person- or episode-level validity assessment is sufficient. Going forward, additional studies that either rule out potential confounding (which includes the risk-adjustment model) or describe features of potential mechanisms will strengthen causal claims.

The risk adjustment methods used are appropriate and demonstrate variation in the prevalence of risk factors across measured entities, contribute to unique variation in the outcome, and show the impact of risk adjustment for providers at high or low extremes of risk. The model performance is acceptable.

Equity: Met

Rationale: The measure sufficiently assesses equity in health care delivery and outcomes, providing crucial insights into how different populations are affected by current practices. The methodology and empirical testing are robust, ensuring that the measure can effectively contribute to ongoing efforts to address and reduce health care inequities.

Use & Usability: Met

Rationale: The measure is actively used in at least one accountability application, with a clear feedback approach that allows for continuous updates based on stakeholder feedback. The specific actions accountable entities can take to improve measure results are described within the logic model and evidence of measure importance.

Public Comment

This measure did not receive any comments during the public comment period.

Advisory Group Feedback

Feedback/Questions	Summary of Developer Response
Similar Feedback to CBE #4630: Advisory Group members had similar questions and feedback surrounding gaming, importance, equity, use and usability, and title as noted within CBE #4630.	Please see response for #4630.
Time Period: An Advisory Group member asked why a 12-month time period was selected. One Advisory Group member added that while the 12-month period allows the measure to capture smaller providers, the actionability of the data is minimalized with the length of time to reporting and, therefore, less improvement happens.	They aggregate data for a year, and the public information is reported on a quarterly basis. This means that every quarter, a 12-month average is shown. This method allows for reporting on a maximum number of facilities. On the actionability of the data, they noted it was a fair point. CMS aims to provide information to providers, ensuring they receive results before the data is publicly reported.
Social Risk Factors: An Advisory Group member observed that, as home health is more community based and less facility based than the other settings, the expectation would be that social risk factors play a greater role. However, they said the developer's regression model seemed to indicate otherwise.	They found the interplay of factors often had the greatest impact on the risk model, but that does not mean a factor was not important or had an impact on its own. One of the biggest factors they observed was whether an individual came to the home health agency from post-acute care or the community.

Key Themes from Advisory Group Feedback and Staff Assessments

Discussion Category	Key Themes	Source of Comment	Summary of Comments
Supportive	Importance	Advisory Group	A few Advisory Group members were pleased to see these measures brought forward and expressed that their cross-setting nature would be beneficial.
Dissenting	Time Period	Advisory Group	The Advisory Group discussed whether the 12-month period was appropriate for the measure, with one Advisory Group member pointing out that such a lengthy time frame made improvement more difficult for clinicians to implement.
	Feasibility	Staff Assessment	In considering the people necessary to implement the measure, the Recommendation Group should assess whether the imputation methodology used in the measure calculation places an undue burden on providers.
	Gaming	Advisory Group	A few Advisory Group members discussed whether the function score could be easily manipulated so that entities could appear to perform better on the measure.

CBE #3645: Hospice Visits in the Last Days of Life [Abt Global/CMS]

Specifications

Measure Description: *The proportion of hospice patients who received hospice visits from a Registered Nurse or Medical Social Worker (non-telephonically) associated with the measured hospice entity during at least two of the final three days of life.*

Staff Preliminary Assessment Rating

Importance: Met

Rationale: This maintenance measure meets all criteria for "Met" due to its well-graded evidence base, clear business case, documented performance gap, and well-articulated logic model.

Feasibility: Met

Rationale: This maintenance measure meets all criteria for "Met" due to its feasibility assessment, clear and implementable data collection strategy, and transparent handling of licensing and fees, ensuring practical implementation within the health care system.

Reliability: Met

Rationale: The results demonstrate sufficient reliability at the accountable-entity level.

Validity: Met

Rationale: Association studies with the measure focus and related outcomes constitute a moderate demonstration of validity. Going forward, additional studies that either rule out potential confounding or describe features of potential mechanisms will strengthen causal claims.

Equity: Met

Rationale: The measure sufficiently assesses equity in health care delivery and outcomes, providing crucial insights into how different populations are affected by current practices.

Use & Usability: Met

Rationale: This maintenance measure is actively used in at least one accountability application, with a feedback approach that allows for continuous updates based on stakeholder feedback. The measure also demonstrates a positive trend in performance results, affirming its ongoing usability.

Public Comment

Number of Comments Received During the Public Comment Period: 2

Comments and their responses from measure developers can be found on the measure page under the “Comments” tab (*Figure 2*).

Advisory Group Feedback

Feedback/Questions	Summary of Developer Response
Importance: Several Advisory Group members, particularly patient participants, expressed that they thought this was an important measure that drives improvements. One member said they were glad this measure supports patients and caregivers.	The developer did not specifically address this comment.
Missing Individuals: An Advisory Group member asked if individuals were potentially being missed by the measure, particularly during changes in care when palliative care ceases and hospice care commences.	The focus on home care, and not inpatient care, is to address instances when an individual is at highest risk of not being seen for a hospice visit. If someone is inpatient, they likely are surrounded by staff.
Holistic Care: Several Advisory Group members expressed that they thought the measure was limited in what it considered a “visit.” A few members said that they wished chaplains or other spiritual care providers were included. One Advisory Group member highlighted that chaplains have a non-billable code. Another Advisory Group member highlighted that licensed counselors and Licensed Marriage Family Therapists (LMFTs) are now licensed to provide care and should be included in the measure as well.	As this is a claims-based measure, they are limited by what information appears on claims, but they can also look at other disciplines for testing. They would look into the non-billable chaplain codes to see if that is something that has shown up on hospice claims.
Timeframe: An Advisory Group member said the measure might be more useful if it looked at 72 hours rather than a calendar day, highlighting that if a patient dies at 2 AM that would be counted as day 1.	The measure looks at 2 of the last 3 days of life. They are limited by how the claims data are captured.
Cultural Considerations: A few Advisory Group members commented on how some cultures, and particularly the Black and African-American communities, are hesitant to engage in hospice because it is seen as “giving up” on family responsibility. They asked how this measure might inform or account for those cultural differences.	They are eager for suggestions on how to improve when questions are related to cultural considerations. Because the measure is claims based, they are unable to tell why a hospice visit may have not happened when one does not occur.
Improvements: An Advisory Group member asked for more information on how this measure has improved processes.	The measure and its improvements are closely aligned with payment incentive, as CMS pays additional for hospice visits.

Feedback/Questions	Summary of Developer Response
Streamlining Measures: An Advisory Group member pointed out that the developer has two measures related to end-of-life visits. They pointed out that the similarity can cause confusion in public reporting.	The measures are part of a broader index method. They will pass this information along to the measure steward.

Key Themes from Advisory Group Feedback and Staff Assessments

Discussion Category	Key Themes	Source of Comment	Summary of Comments
Supportive	Importance	Advisory Group	The Advisory Group, and particularly the patient participants, said that this was an important measure that can drive improvement and support patients and caregivers.
Dissenting	Holistic Care	Advisory Group; Public Comment	Several Advisory Group members stated that a limitation of the measure was its inability to capture other disciplines that may provide care at end of life, including chaplains and social workers. Public comments underscored this same limitation, noting that the measure should be revised to include a broader range of visit types, allow virtual visits, and consider patient and family wishes to better reflect quality care. In addition, the measure only focuses on the frequency of Registered Nurse (RN) and Medical Social Worker (MSW) visits rather than the quality of care or patient and family preferences.
	Usability	Public Comment	The National Alliance of Care expressed that the measure’s requirement to identify the “final days” of life is imprecise, and its implementation has shown minimal improvement in hospice performance, especially during the COVID-19 pandemic.